

Recherche d'Informations L3 (ISIL)

TD1

Q1 : Quelle est la condition pour qu'un corpus soit significatif ?

Q2 : Y a-t-il une différence entre un système de base de données et un système de recherche d'information ?

Q3 : Un terme qui apparaît dans tous les documents d'un corpus est-il discriminant et pourquoi ?

Q4 : Quels sont les différents acteurs d'une recherche d'informations ?

Q5 : donnez un exemple de chaque système en précisant la partie destinée à une machine et celle envoyée à l'utilisateur.

1-Une base de données structurée.

2-Une base de données semi-structurée

3-Une base de données non structurée.

Q6. Pourquoi l'indexation est-elle une étape essentielle dans un système

Q7. Expliquez la notion de pertinence dans un système de recherche d'informations.

Exercice 2 Soit les informations suivantes contenues dans **le tableau 1** concernant un corpus C pour les requêtes indiquées. Calculer les taux de précision et de rappel du système à chaque réponse et remplir le tableau ci-dessus. Supposons que pour les requêtes indiquées, le nombre total de documents pertinents dans l'ensemble de tous les documents est 14.

Exercice 3 :

Soit le **tableau 2** résumant les résultats obtenus avec un système de recherche.

□ Donner la précision et le rappel.

□ Quelle méthode donne toujours un rappel maximal ? une précision maximale?

Calculer la F-mesure.

Exercice 4

Considérons les documents suivants :

Doc1 : "On mesure l'efficacité d'un SRI par les paramètres précision et rappel."

Doc2 : " Les modèles de recherche les plus efficaces sont le modèle de langage et le modèle vectoriel."

Doc3 : "L'efficacité de la recherche est mesurée par la précision moyenne."

Doc4 : " Apparition du modèle booléen et l'élaboration de petites expérimentations sur des petites collections de documents. "

Doc5 : " L'indexation est un processus permettant de construire un ensemble d'éléments clés permettant de caractériser le contenu d'un document et de retrouver ce document en réponse à une requête."

1. Réaliser l'indexation de ces documents en précisant à chaque phase le résultat obtenu.
2. Construisez le fichier inverse en considérant uniquement D2 et D3.

Rang	n°doc	Pertinent	Rappel	Précision
1	324	X		
2	589	X		
3	528	X		
4	590	X		
5	986	X		
6	592	X		
7	899	X		
8	988	X		
9	578			
10	985			
11	537	X		
12	591	X		
13	772	X		
14	990			

Tableau1 (Exercice 2)

	Pertinent	Non pertinent
Sélectionné	50	10
Non sélectionné	30	120

Tableau 2 (Exercice 3)

